



# **2-Day Course – Spatial Modeling with Geostatistics**

**Prof. Michael J. Pyrcz, Ph.D., P.Eng.  
Associate Professor**

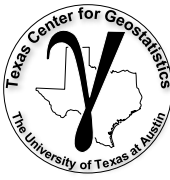
**Hildebrand Department of Petroleum & Geosystems Engineering  
University of Texas at Austin**

**Bureau of Economic Geology, Jackson School of Geosciences  
University of Texas at Austin**

**“In two days, what a geoscientists needs to know about geostatistics, and  
workflows to get you started with applying geostatistics to impact your work.”**

# Spatial Modeling with Geostatistics

## Prerequisites



Lecture outline . . .

- Who am I?
- Class Objectives
- Class Strategy
- Essential Pre-work
  - Installation, set up

**Prerequisites**

Introduction

Probability Theory

Representative Sampling

Spatial Data Analysis

Spatial Estimation

Stochastic Simulation

Uncertainty Management

Machine Learning

# Class Description. . .

## Who am I?

- new this fall to UT PGE.
- over 17 years of experience in consulting, teaching and industrial R&D in statistical modeling, reservoir modeling and uncertainty characterization.
- associate editor with Computers and Geosciences, nominee to International Assoc. of Mathematical Geosciences committee.
- member of scientific committee for Geostatistical Congress 2016.
- author of the textbook “Geostatistical Reservoir Modeling” and > 40 peer reviewed publications, patents etc.
- “I want to give you a competitive edge in your careers with geostatistics.”





# Class Objectives

Teach theory and practical methods for geostatistics.

Communicate:

- the benefits and uses of geostatistics,
- the common spatial and uncertainty modeling workflows,
- how to better integrate their domain knowledge into the geostatistical model.

Provide knowledge and resources to start geoscientists building their own workflows.

Initial experience with workflow construction with open source



# Class Strategy

A combination of lecture, demonstration and hands-on

## Lectures

- Provide the fundamental theory with a focus on practice

## Demonstrations

- Illustrate the use of geostatistics with open source to solve practical problems
- I will use Python / GSLIB, Excel, and R workflows that are available to students

## Hands-on

- Experiential learning with R / R Studio and “gstat” package

# Hands-On

We will conduct hands-on in R, because it is easiest for getting started, and has robust packages for geostatistics and data analytics

It will require students to complete the following before the class.

1. Install R from one of the mirror sites (e.g. <http://cran.wustl.edu/>)
2. Install R Studio from <https://www.rstudio.com/products/rstudio/download/>. The free version is fine.

# Hands-On

## 3. Open R Studio

## Step Through R Code

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

real1 x PCA\_demo.Rmd x variogram\_demo.R x kriging\_demo.R x simulation\_demo.R x PCA\_demo.R x simulation\_Balhoff.R x DT\_demo.Rmd x **DT\_demo.R** x

```

18 # no functions required in this demonstration
19
20 # Set the working directory, I always like to do this so I don't lose files and to simplify subsequent read and writes
21 setwd("c:/PGE337/DT") # choose your local working directory
22
23 # Read the data table from a comma delimited file - data on GitHub/GeostatsGuy/GeoDataSets
24 mydata = read.csv("unconv_MV_V2.csv") # read in comma delimited data file
25
26 # Let's visualize the first several rows of our data so we can make sure we successfully loaded it
27 head(mydata) # show the first several rows of a data table in the console
28
29 # Check out the summary statistics for each column
30 summary(mydata) # summary statistics for the multivariate data file
31
32 # Calculate the correlation matrix
33 mydata_noindex <- mydata[,2:length(mydata)] # remove first column with the well index
34 cor_matrix <- round(cor(mydata_noindex),2) # calculate a mxm matrix with the correlation coefficients
35 cor_matrix
36
37 # Let's use the corplot package to make a very nice correlation matrix visualization
38 corplot(cor_matrix, method = "circle") # graphical correlation matrix plot
39
40 # Now let's view the scatterplot matrices from the lattice Package
41 splom(mydata[c(2,3,4,5,6,7,8)],col=rgb(0,0,0,0.50,maxColorValue=255), pch=19,main = "Unconventional Dataset")
42 # This dataset has variables from 1,000 unconventional wells including well average porosity, log transform
43 # of permeability (to linearize the relationships with other variables), acoustic impedance (kg/m2s*10^6), brittleness ratio (%),
44 # total organic carbon (%), vitrinite reflectance (%), and production (MCFPD)
45
46 # Let's start simple with a trivariate (3 variable) problem
47 mydata_por <- data.frame(mydata[1:1000,2]) # extract and rename 3 features from the original dataframe
48 colnames(mydata_por) <- "Por"
49 mydata_brittle <- data.frame(mydata[1:1000,5])
50 colnames(mydata_brittle) <- "Brittle"
51 mydata_prod <- data.frame(mydata[1:1000,9])
52

```

48T (Top Level) R Script

Console C:/PGE337/DT/

```

1st Qu.: 2120.961
Median : 2991.763
Mean : 3237.154
3rd Qu.: 4105.623
Max. : 12568.644

> mydata_noindex <- mydata[,2:length(mydata)] # remove the first column with the well index
> cor_matrix <- round(cor(mydata_noindex),2) # calculate a mxm matrix with the correlation coefficients
> cor_matrix


```

Por LogPerm AI Brittle TOC VR Production Prod2Scaled

	Por	LogPerm	AI	Brittle	TOC	VR	Production	Prod2Scaled
Por	1.00	0.81	-0.51	-0.25	0.71	0.08	0.69	0.86
LogPerm	0.81	1.00	-0.32	-0.14	0.51	0.05	0.57	0.70
AI	-0.51	-0.32	1.00	0.16	-0.55	0.49	-0.33	0.54
Brittle	-0.25	-0.14	0.16	1.00	-0.24	0.29	-0.08	0.13
TOC	0.71	0.51	-0.55	-0.24	1.00	0.31	0.50	0.62
VR	0.08	0.05	0.49	0.29	0.31	1.00	0.14	0.13
Production	0.69	0.57	-0.33	-0.08	0.50	0.14	1.00	0.96
Prod2Scaled	0.86	0.70	-0.42	-0.15	0.62	0.13	0.96	1.00

```

> corplot(cor_matrix, method = "circle") # graphical correlation matrix plot
> splom(mydata[c(2,3,4,5,6,7,8)],col=rgb(0,0,0,0.50,maxColorValue=255), pch=19,main = "unconventional Dataset")
> mydata_por <- data.frame(mydata[1:1000,2]) # extract and rename 3 features from the original dataframe
>

```

Text Output

Environment History

Global Environment

Data

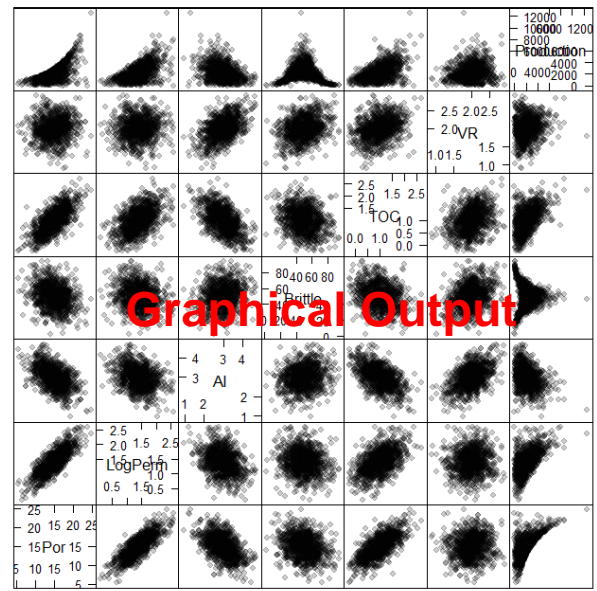
- cor\_matrix num [1:8, 1:8] 1 0.81 -0.51 -0.25 0.71 0.08 0.69 0.86 0.86
- loaded\_por index 5 obs. of 2 variables
- mydata data frame with 8 variables
- mydata\_noindex 1000 obs. of 8 variables
- mydata\_por 1000 obs. of 1 variable

Variables / Objects

Files Plots Packages Help Viewer

Zoom Export Publish

Unconventional Dataset



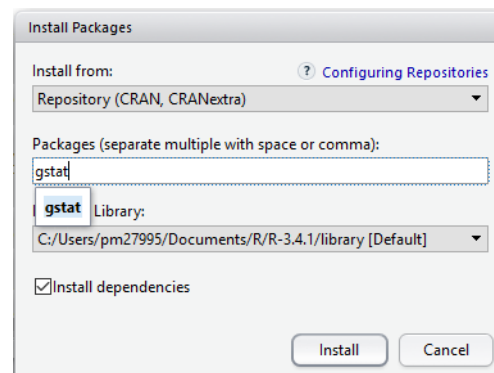
Scatter Plot Matrix

# Hands-On

## 4. Install the following packages

- `gstat` geostatistics package by Edzer Pebesma
- `sp` adds spatial to DataFrames
- `plyr` manipulating data
- `ggplot2` plotting
- `fields` plotting regular grid models
- `lattice` matrix scatter plots
- `corrplot` correlation plots
- `tree` decision trees

To install packages go to Tools/ Install Packages...



Enter name of package and select install.

Ignore R Version warnings.



# Hands-On

5. Open provided R code, kriging\_demo.R

[https://github.com/GeostatsGuy/geostatsr/blob/master/kriging\\_demo.html](https://github.com/GeostatsGuy/geostatsr/blob/master/kriging_demo.html)

6. Change the working directory

```
# Set the working directory, I always like to do this so I don't lose files and to simplify subsequent read and writes  
setwd("C:/PGE337")
```

- Change from C:/PGE337 to a folder of your choice on your computer
7. Download these datasets from GitHub and put them in your working folder.
- **2D\_MV\_200Wells.csv**
    - [https://github.com/GeostatsGuy/GeoDataSets/blob/master/2D\\_MV\\_200wells.csv](https://github.com/GeostatsGuy/GeoDataSets/blob/master/2D_MV_200wells.csv)
  - **unconv\_MV\_v2.csv**
    - [https://github.com/GeostatsGuy/GeoDataSets/blob/master/unconv\\_MV\\_v2.csv](https://github.com/GeostatsGuy/GeoDataSets/blob/master/unconv_MV_v2.csv)
  - **unconv\_MV\_v3.csv**
    - [https://github.com/GeostatsGuy/GeoDataSets/blob/master/unconv\\_MV\\_v3.csv](https://github.com/GeostatsGuy/GeoDataSets/blob/master/unconv_MV_v3.csv)

# Hands-On

8. Back in R Studio window, place the cursor at the top of the code and step through the code with the “run” button indicated on slide 6.
9. Watch the text and graphical output. Check text output for errors. Warnings are fine.
  - A couple of the numerical methods may take 15 to 30 seconds to complete so be patient.
  - Resist the temptation to “machine gun” the run button as this may cause a crash of R studio.
10. If you get to the end of the code file then you should be set up and good-to-go for the hands-on sections.



# Hands-On

- If you are not able to complete this set up, then you can pair up with another student to work together on the hands on sections.

# What did you just learn?

## Lecture outline . . .

- Who am I?
- Class Objectives
- Class Strategy
- Prerequisites

**Prerequisites**

**Introduction**

**Probability Theory**

**Representative Sampling**

**Spatial Data Analysis**

**Spatial Estimation**

**Stochastic Simulation**

**Uncertainty Management**

**Machine Learning**